

Lodtrækningsforsøg og andre eksperimentelle design i Danmark

**Mette Deding, Forskningsleder,
SFI – Det Nationale Forskningscenter for Velfærd,
mcd@sfi.dk**

**Mads Andersen Høg, Ledelseskonsulent,
SFI – Det Nationale Forskningscenter for Velfærd,
mah@sfi.dk**

Denne artikel belyser en række strukturelle, etiske og kulturelle udfordringer i forhold til at gennemføre lodtrækningsforsøg på social- og velfærdsområdet i Danmark. Lodtrækningsforsøg er teoretisk set den bedste måde at bestemme kausal effekt af en indsats, og den viden er afgørende, hvis man vil styrke det sociale sikkerhedsnet og styre socialpolitikken, så man får størst mulig effekt for pengene. I praksis har det imidlertid vist sig, at der er en række udfordringer i forhold til at udføre lodtrækningsforsøg i Danmark, bl.a. i forhold til at etablere en kontrolgruppe, i samarbejdet mellem forskning og praksis og i forhold til hvordan vi generelt designer sociale indsatser i Danmark. Der findes også andre (quasi-eksperimentelle) metoder, der kan estimere effekter i de tilfælde, hvor det ikke kan lade sig gøre at lave lodtrækningsforsøg, men de indebærer en række metodiske ulemper, som kan gøre estimatet af effekter usikkert.

»Virker det?« »Hvad er effekten?« De spørgsmål er i de senere år kommet til at fylde mere og mere i den danske politiske debat. På social- og velfærdsområdet er lodtrækningsforsøg så småt ved at blive en fast del af værktøjsskassen, men der er stadig en række udfordringer af såvel strukturel, som etisk og kulturel karakter. Derudover bliver forskellige andre økonometriske metoder brugt til at forsøge at svare på den store efterspørgsel på viden om kausal effekt. Men også her er der udfordringer.

Årsager til fokus på effekt

Der er flere årsager til, at der er kommet stadig større fokus på spørgsmålet om effekter af forskellige policy-tiltag både fra politikerside og fra praktikerside. På det social- og velfærdspolitiske område – som er det område, vi på SFI arbejder inden for – er én af årsagerne en stigende erkendelse af, at det universelle sikkerhedsnet af ydelser og indsatser som den danske velfærdsstat består af, ikke fanger alle. Afhængigt af, hvilket område man kigger på, så er der rimelig konstant ca. 10-20 pct. af befolkningen, der ligger i bunden og som vi har svært ved at få til at flytte sig med traditionelle midler (et eksempel er børneområdet, hvor SFI's børneforløbsundersøgelse af børn født i 1995 løbende har dokumenteret, at mens de fleste trives og har det godt, er der vedvarende en restgruppe på 10-20 pct. med forskellige vanskeligheder: fx Christensen (2000), Christensen (2004), Mattson o.a. (2008) Ottosen (2012)). Af forskellige årsager (menneskelige, politiske, økonomiske) vil vi gerne gøre noget for denne »restgruppe«, hvad enten det er udsatte børn, flygtninge eller indvandrere, eller borgere på kontanthjælp langt fra arbejdsmarkedet. Det er ikke, fordi der ikke bliver forsøgt gjort noget for disse grupper – i form af forskellige indsatser og projekter – men hvis vi for alvor vil hjælpe og rykke, er

det nødvendigt at finde målrettede indsatser, som har en målbar effekt.

En anden vigtig årsag er, at det er nødvendigt at sikre sig, at de indsatser, man laver, ikke har negativ effekt. På socialområdet i Danmark har der ikke været nogen synderlig udbredt kultur om at måle effekten af det, man laver. Alle indsatser er selvfølgelig sat i verden med intentionen om at gøre en positiv forskel for målgruppen, men vi ved fra internationale erfaringer, at på trods af gode intentioner, så findes der faktisk indsatser, som har den modsatte effekt af, hvad man ønskede.

Det klassiske eksempel er »scared straight«, hvor man satte ungdomskriminelle ind i voksenfængsler ud fra den hypotese, at det ville virke så afskrækkende på de unge at møde de voksne kriminelle, at de derefter ville holde sig langt væk fra kriminalitet. Grundige evalueringer – bl.a. ved hjælp af lodtrækningsforsøg – viste dog, at effekten var den modsatte: De ungdomskriminelle blev altså mere kriminelle af indsatsen, sandsynligvis fordi de opfattede de voksen-kriminelle som positive rollemodeller (Petrosino o.a., 2013).

Men også andre indsatser kan have en utilsigtet effekt. Pointen er, at det vi går rundt og tror, ikke nødvendigvis viser sig at være rigtigt, og at det derfor er nødvendigt at undersøge. Mange klassiske evalueringer anvender fx forskellige tilfredshedsvurderinger, og disse kan selvfølgelig være relevante for mange forskellige spørgsmål om fx indretning og implementering af en given indsats. Men »tilfredshed« har ikke noget at gøre med effekt. Det er derfor nødvendigt at kunne skelne mellem eksempelvis brugernes og fagpersonernes umiddelbare vurdering af en indsats, og så den reelle effekt. Det kan være svært at overskue de reelle og måske mere langsigtede virkninger, når man med de bedste intentioner er engageret i at gøre en forskel.

En tredje og mere generel årsag til det øge-

de fokus på effekt er naturligvis den stadig større kamp om de begrænsede ressourcer og behov for prioritering, der er i den offentlige sektor. Viden om effekt fjerner ikke behovet for prioritering, men det giver et mere sagligt funderet beslutningsgrundlag, når vi ved, hvilke indsatser, der rent faktisk flytter noget i den rigtige retning – og også, hvilke indsatser der giver den størst mulige effekt i forhold til omkostningerne.

Lodtrækningsforsøget

Det store fokus på effekter har i en vis grad betydet, at begrebet »effekt« er blevet lidt udvandet. »Effekt« og »hvad virker?« har gennem de seneste år været de helt store buzz-words på socialområdet, og udtrykkene bliver brugt lidt i flæng. Derfor er det vigtigt at understrege, at når vi her taler om effekt, så mener vi *kausal effekt*. Vores fokus er altså undersøgelser, der kan svare på det kontrafaktiske spørgsmål: Hvordan går det i gennemsnit en gruppe af personer, der modtager en indsats, i forhold til hvordan det ville være gået, hvis de *ikke* havde modtaget en indsats?

Fordi det kontrafaktiske spørgsmål i sagens natur ikke kan besvares (samme gruppe af personer kan ikke modtage en indsats samtidig med at de ikke modtager indsatsen), er hjørnестenen i effektmåling at etablere en kontrolgruppe (der ikke modtager indsatsen), der er sammenlignelig med en indsatsgruppe (der modtager indsatsen). Teoretisk set er lodtrækningsforsøget (RCTs) det stærkeste design i forhold til at sikre sammenligneligheden mellem indsats- og kontrolgruppe. Fordi randomiseringen i lodtrækningsforsøget sker på en samlet population, *før* indsatsen går i gang, er de to grupper som udgangspunkt statistisk ens i forhold til alle baggrundskarakteristika – både de observerede og ikke-observerede. Endvidere vil begge grupper blive udsat for samme påvirkninger udefra i projektperioden (såsom generelle lovændringer der kan påvirke borgerne uafhængigt af indsatsen). Den eneste statistiske forskel på

indsats- og kontrolgrupperne er således, at den ene gruppe modtager indsatsen, mens den anden ikke gør. Hvis forsøget ellers er udført ordentligt, er det derfor for så vidt ret simpelt at finde effekten, idet man bare kan sammenligne, hvordan det går indsats- og kontrolgrupperne efterfølgende i forhold til de centrale outcomes, man er interesseret i (Torgerson og Torgerson, 2008).

Som nævnt er lodtrækningsforsøget i teorien det metodisk stærkeste. Forbeholdet »i teorien« dækker selvfølgelig over, at det i praksis ikke altid er nemt at gennemføre et lodtrækningsforsøg (Hansen og Pontoppidan, 2014). Hele identifikationsstrategien ligger i randomiseringen, og hvis den af den ene eller anden årsag glipper, fejler identifikationen af den kausale effekt tilsvarende. Et lodtrækningsforsøg kræver tæt kontrol med, at randomiseringen sker, som den skal – altså helt tilfældigt. Der kan være mange årsager til, at randomiseringen fejler. Det kan være, at velmenende sagsbehandlere (eller forskere) tager enkelte sager i egen hånd og tildeler indsats ud fra deres egen faglige vurdering i stedet for via randomisering (vi skal senere vende tilbage til denne problemstilling). En anden problemstilling kan være, at kontrolgruppen ikke vil deltage i forsøget, hvis bare de »bare« er kontrol, eller at de af sig selv opsøger den indsats, som de ikke var udset til at få. Derudover er der andre kendte problematikker med lodtrækningsforsøg, som fx Hawthorne-effekten, der påpeger, at selve det at deltage i et forsøg kan betyde noget for deltagere og dermed deres udbytte af indsatsen (Torgerson og Torgerson, 2008).

Disse problemstillinger gør, at man både ved udførelse og ved vurdering af lodtrækningsforsøg skal være mere end almindeligt opmærksom på, hvordan selve randomiseringen er foregået (Higgins og Green, 2008). Man taler i den forbindelse fx om »blinding«, altså at det holdes skjult for både deltagere (der enten er i en indsats- eller i en kontrolgruppe),

praktikere (der er inddraget i visitation og udførelse af de konkrete indsatser) og forskere (der skal evaluere forsøget), hvem der er i den ene hhv. den anden gruppe. Hvis man ikke er vidende om, hvem der er i hvilken gruppe, er det selvklart svært at manipulere med grupperne, mens der findes talrige eksempler på, at forsøg uden blinding er blevet manipulerede bevidst eller ubevidst, hvilket selvfølgelig ødelægger forsøgets styrke.

Lodtrækningsforsøget som metode er meget udbredt på tværs af fagområder (og er fx langt mere udbredt på sundhedsområdet end på socialområdet), og heraf følger faktisk en af de store fordele ved metoden. For selvom der er en række metodiske udfordringer ved lodtrækningsforsøget – som der i øvrigt er med alle metoder – så er de i denne forbindelse systematisk dokumenteret, og der findes en række internationalt anerkendte værktøjer, der kan bruges til kvalitetsvurdering af det enkelte forsøg; eksempler er CONSORT (www.consort-statement.org/) og PRISMA (www.prisma-statement.org/). Endvidere er der en anden faglig kultur omkring afrapportering af lodtrækningsforsøg, som gør, at det er nemt for andre forskere at granske og vurdere forsøgets resultater. Fx kræver en forskningsfaglig solid afrapportering af et lodtrækningsforsøg, at forskerne har publiceret en protokol med beskrivelse af randomisering og analyseplan, *inden* analysearbejdet går i gang. Det sætter andre forskere i stand til efterfølgende at vurdere, om analyseplanen er holdt og om forsøget er validt i forhold til at estimere en kausal effekt (Higgins og Green, 2008).

Strukturelle udfordringer

I Danmark er lodtrækningsforsøg som metode på det social- og velfærdspolitiske område ikke særligt udbredte og af nyere dato. Det er i den forbindelse interessant at bemærke, at når det gælder lodtrækningsforsøg på sundhedsområdet, så gennemfører Danmark relativt til indbyggertal rigtig mange lodtræk-

ningsforsøg, så det er ikke, fordi danskere per definition er modstandere af lodtrækning (<https://clinicaltrials.gov/ct2/home>). Men den langsomme fremdrift på social- og velfærdsområdet kan skyldes strukturelle udfordringer på området eller barrierer af mere etisk og kulturel karakter, hvilket vi vil komme ind på i det følgende.

Når vi kigger ud i den store verden, er det tydeligt, at USA har haft en helt anden tradition for at lave lodtrækningsforsøg af sociale indsatser sammenlignet med Danmark og Europa i det hele taget. En af årsagerne til den forskel er selvfølgelig evalueringstradition, men det hænger også sammen med måden, vi har indrettet vores velfærdssystem på. Hvis man skal måle en effekt ved hjælp af et lodtrækningsforsøg, kræver det blandt andet, at man kan afgrænse en målgruppe for indsatsen, at nogle, men ikke alle får indsatsen, og at man kan afgrænse og beskrive indsatsen så detaljeret, at andre efterfølgende kan bruge resultatet. Det danske velfærdssamfund bygger langt hen af vejen på universelle og meget brede indsatser, hvilket helt strukturelt gør det vanskeligt at måle effekter via lodtrækningsforsøg (Konnerup, 2011).

Den nemmeste måde at gennemføre et lodtrækningsforsøg på er, hvis man kan tilbyde en ekstra indsats til indsatsgruppen, mens kontrolgruppen ikke bliver stillet dårligere end ellers. Det kunne fx være, hvis man vil afprøve en ny social metode. Så bliver den nye metode tilbudt til indsatsgruppen, mens kontrolgruppen får »treatment-as-usual«, altså den behandling, de ville have fået i fravær af den nye metode. På den måde bliver ingen stillet ringere som følge af forsøget. Fx blev effekten af et nyt træningsprogram til ældre i Gladsaxe Kommune undersøgt ved at lade en tilfældig udvalgt gruppe af ældre deltage i det nye forløb, mens en anden gruppe af ældre fik den hidtidige behandling (Termansen og Sonne-Schmidt, 2012).

Men der ligger en strukturel udfordring i at lave forsøg, der indebærer nedskæringer, fordi man da i forsøgsperioden måske kommer til at stille nogen dårligere end ellers. Et eksempel er et forsøg, der blev planlagt for at vurdere effekten af en mere effektiviseret sundhedsplejerskeordning, hvor sundhedsplejen blev samlet i centre, hvor mødre og børn kan møde op og bliver målt og vejlet, frem for de hidtidige hjemmebesøg. Det er oplagt, at den nye ordning er billigere end den gamle, men ikke så oplagt, om det vil have betydning for fx mødres og børns trivsel. Og ligeledes ikke oplagt, hvor stor omkostningen ved en eventuelt negativ effekt var i relation til de sparede udgifter på området. Det blev derfor foreslået at gennemføre et lodtrækningsforsøg, hvor nogle af de nybagte mødre blev randomiseret til stadig at modtage hjemmebesøg af sundhedsplejersken, mens den anden gruppe overgik til at besøge de nye sundhedsplejerskecentre. Dette blev dog opfattet som for problematisk at argumentere for fra politisk side, så forsøget måtte droppes (men den nye ordning blev gennemført).

En anden strukturel udfordring ved lodtrækningsforsøget er det faktum, at Danmark er et lille land. Det betyder til tider, at målgrupperne for specifikke indsatser er ganske små, og det kan gøre det svært at opdele gruppen i to, som jo er forudsætningen for lodtrækningsforsøget. I lodtrækningsforsøg beregner man »styrke«, dvs. hvor stort vores sample af deltagere skal være for at kunne påvise en effekt af en vis forventet størrelse med en given statistisk sandsynlighed. Hvis der er for få deltagere, er problematikken, at et insignifikant resultat selvfølgelig kan skyldes, at der ikke er nogen effekt, men også kan skyldes, at der ikke har været nok deltagere til at påvise en effekt. Hvis den forventede effekt af en given indsats kun er lille, kan man derfor komme i den situation, at der skal rekrutteres urealistisk mange deltagere til forsøget. Igen kan det dog fremhæves som en styrke ved lodtrækningsforsøget som metode, at

problemstillingen træder tydeligt frem allerede ved styrkeberegningen, og altså allerede inden forsøget går i gang, så man ikke efterfølgende risikerer at have brugt tid og kræfter på et projekt, som ikke har den store sandsynlighed for at give brugbare resultater.

Etiske udfordringer

Men ud over disse strukturelle udfordringer i forhold til at lave lodtrækningsforsøg på det sociale område i Danmark er der en række etiske problemstillinger, som man selvfølgelig også må forholde sig til. Kernen i lodtrækningsforsøget er at etablere en forskel mellem indsats- og kontrolgruppe, men netop den strukturerede tilfældige forskel, som lodtrækningen sikrer, møder ofte modstand blandt praktikere og andre ikke-forskere. Det føles forkert at skulle trække lod om, hvilken indsats man skal give til eksempelvis en job-søgende eller en udsat ung. Den umiddelbare reaktion er, at det burde være et fagligt skøn, der skal ligge til grund for indsatsen; ikke tilfældighed. Samtidig kræver lodtrækningsforsøget, at man som eksempelvis socialrådgiver i en periode skal overlade det til andre at bestemme, hvilken indsats en bestemt borger skal have. Fagfolkene skal i en vis grad fralægge sig beslutningskompetencen, og det kan – helt forståeligt – være svært.

Svaret på denne instinktive modstand kan – udover at informere grundigt om baggrunden for forsøget – være at hæve perspektivet fra den enkelte borger til den samlede mængde af sager. For selv om alle borgere i henhold til lovgivningen skal behandles ens, så behandles de de facto uens. På tværs af kommunegrænser og nogle gange også på tværs af medarbejdere iværksættes forskellige typer af indsatser målrettet samme problemstillinger. Og hvis vi reelt ikke kender effekten af den ene frem for den anden indsats, så behandles folk i realiteten forskelligt. Lodtrækningsforsøget er blot en struktureret metode til at udnytte denne eksisterende forskel til at opnå viden om effekter.

En anden etisk udfordring ligger i, hvordan vi informerer de borgere, der deltager i et lodtrækningsforsøg. Udgangspunktet er, at deltagerne altid skal informeres grundigt og give deres samtykke, inden de deltager i et forsøg. Og at de skal have mulighed for at sige fra, hvis de ikke vil deltage i forsøget, uden at det går ud over deres muligheder fx for at få et kommunalt tilbud. Dette er som udgangspunkt arbejdsgangen, hvis der i forbindelse med forsøget skal indsamles forskellige data fra deltagere – fx spørgeskemadata – der ellers ville blive indsamlet. Men nogle gange er det uhensigtsmæssigt at skulle have kontakt med hver enkelt deltager, fx hvis randomiseringen sker på et administrativt niveau, og analysen kun baseres på allerede forekommende datakilder. I sådan et tilfælde kan det ødelægge forsøget at skulle indhente samtykker, og i øvrigt ødelægge et randomiserings-setup med fuld blinding af deltagerne. Men omvendt kan man jo diskutere, hvorvidt borgerne altid har krav på information om at de deltager i et lodtrækningsforsøg. Det kan i øvrigt i den forbindelse bemærkes, at vores erfaring er, at meget få deltagere takker nej til at deltage i et lodtrækningsforsøg, når de bliver informeret om det, mens barrieren er meget større fra systemets side.

Man må ikke undervurdere betydningen af disse udfordringer, for samarbejdet mellem forskning, praksis og borger er helt afgørende, hvis et lodtrækningsforsøg skal gennemføres ordentligt. Det er vigtigt, at de frontmedarbejdere, der skal implementere den konkrete indsats, kan stå inde for designet bag forsøget. Og hvis ikke de udførende led er tro mod designet, kan der hurtigt opstå fejlkilder – bl.a. i randomiseringen som påpeget ovenfor.

En måde, vi forsøger at imødekomme de etiske forbehold på, er ved at informere alle parter grundigt om projektet. Et eksempel er et projekt om skolestøtte til børn i familiepleje, hvor børn anbragt i

plejefamilier er randomiseret i tre grupper – én gruppe modtager indsatsen »forældre som lektiehjælpere«, en anden gruppe modtager indsatsen »LUKoP for skolen«, mens den tredje gruppe er kontrolgruppe. Dette projekt har en hjemmeside, hvor deltagende plejeforældre, skoler og kommuner kan finde information om projektet (www.sfi.dk/skole-projekt). En anden mulighed, der ville hjælpe med at håndtere disse etiske problemstillinger fremadrettet, kunne være at oprette et social-etisk råd, som kan regulere og vejlede i disse spørgsmål. Der findes allerede et sundhedsetisk råd, som bl.a. tager stilling til etiske problemstillinger i lodtrækningsforsøg på det medicinske område. Fordelen ved et social-etisk råd kunne være, at man derved får et uvildigt forum, hvor man åbent kan diskutere de etiske problemstillinger, der naturligvis er, når man trækker lod, men hvor der også kunne være mulighed for at diskutere andre etiske udfordringer forbundet med evaluering.

Kulturelle udfordringer

Et andet område, som skaber udfordringer for at lave lodtrækningsforsøg i Danmark, er den kultur, vi har for at lave sociale indsatser. På mange måder har vi hyldet princippet om at »lade de 1000 blomster blomstre«. Sagt lidt firkantet har hver kommune eller hver institution udviklet deres egne måder at arbejde på (»metoder«), og den lokale forankring og metodefrihed har været vægtet højere end kravet om metodisk gennemsigtighed og stringens. En bærende tanke har været, at indsatser skal udtænkes og udføres mere eller mindre lokalt.

Der er dog tegn på, at tendensen er begyndt at gå i den modsatte retning. Den stigende efterspørgsel efter viden om »indsatser, der virker« følges i stigende grad af en diskussion om, at »metodefrihed« ikke er det samme som »frihed for metode«. Samme metode kan selvkært ikke anvendes i forhold til alle borgere/alle problemstillinger, men diskussionen begynder at gå på, at metodefrihed må

handle om, at man på et oplyst grundlag træffer beslutning om, hvilken metode der passer bedst i det givne tilfælde, og ikke at man helt frit vælger, hvordan man arbejder. Og en del af dette beslutningsgrundlag er viden om effekter.

Der er dog stadig udfordringer i forhold til tankegangen omkring lodtrækningsforsøget. For hvis man vil lave lodtrækningsforsøg, kræver det både en vis volumen (»nok« deltagere) og vis grad af standardisering (altså at det tilnærmelsesvis er samme indsats der undersøges på tværs af deltagerne). Dette kan gøre det svært at skabe ejerskab til forsøget lokalt, hvor det skal implementeres. Vores erfaring er således, at det er altafgørende med ledelsens opbakning til projektet, for at de kulturelle udfordringer kan overvindes.

En lidt omvendt, men stadig betydelig kulturel udfordring med lodtrækningsforsøg på det sociale område, er systemets utålmodighed. Hvis man fra politisk side har valgt at sætte fokus på effekt, så vil man gerne have svarene hurtigt. Dette er der en række problemer med. For det første kan det tage tid, før det er meningsfuldt at vurdere effekten – tunge sociale problemer bliver typisk ikke fjernet med et snuptag, så indsatserne kan være langvarige, og derudover vil man gerne vide om en eventuel effekt holder længere end bare den periode, hvor selve indsatsen foregår. Et klassisk eksempel er misbrugsbehandling, som nærmest altid virker – misbrugere har et lavere misbrug når de afslutter behandlingen end da de startede behandlingen. Men det er til gengæld meget varierende, hvor meget denne effekt holder sig over tid (Tanner-Schmidt o.a., 2013).

For det andet skal indsatsen være moden til evaluering, før man laver et lodtrækningsforsøg i fuld skala. Hvis der fx er tale om en nyudviklet indsats, der først skal justeres og implementeres, så kan det i sig selv tage flere år. Sætter man et stort lodtrækningsforsøg i

gang for tidligt, risikerer man derfor at måle effekten af en indsats under implementering fremfor den endelige indsats, hvilket var det gentlige mål. Omvendt kan lodtrækningsforsøget med fordel bruges til at pilotteste nye indsatser i udviklingsfasen i et mere kontrolleret setup, før man beslutter sig for om indsatsen eventuelt skal rulles bredere ud – et såkaldt *efficacy*-studie, hvor man afprøver om indsatsen *kan* have en positiv effekt i modsætning til et *effectiveness*-studie, der undersøger effekt når indsatsen anvendes i den virkelige verden (Phil o.a., 2005).

Endnu en kulturel barriere, som skal overkommes, før vi får flere lodtrækningsforsøg i Danmark, er en relativt udbredt forestilling om, at lodtrækningsforsøg er meget dyre og meget omstændelige at gennemføre. Her er det vigtigt at påpege, at det ikke er selve randomiseringen, der er den store udgift. Som oftest er de lodtrækningsforsøg, man i Danmark har sat i gang på socialområdet, store projekter, hvor man i relativt stor skala implementerer en eller flere indsatser, som der kan trækkes lod imellem. Men det er nødvendigt at understrege, at det er implementeringsdelen, der er den store udgift her. Og hvis man vil lave en ny indsats, så skal den jo implementeres under alle omstændigheder. For eksempel vedtog Horsens Kommune for nogle år siden at implementere et IT-støtte-værktøj i alle kommunens folkeskoler. Indsatsen var således allerede besluttet, men fordi vi fik mulighed for at randomisere, i hvilken rækkefølge skolerne skulle implementere indsatsen, blev det muligt for meget beskedne midler at evaluere effekten af det nye tiltag (Christensen o.a., 2014).

Som nævnt er det vigtigt, at man nøje overvejer, hvornår i forløbet man implementerer, og hvornår man måler via lodtrækning. Det afgørende er, at man har gennemtænkt forløbet, inden man går i gang med implementeringen. Hvis man ikke har overvejet muligheder for randomisering, inden man implementerer,

kan man meget hurtigt komme i den situation, at man ikke er i stand til at måle effekten af indsatsen ved hjælp af den stærkeste metode.

Evaluerings af en indsats, efter at den er implementeret, er umiddelbart betydeligt billigere, end et stort lodtrækningsprojekt, hvor implementering indgår. Men desværre betyder evaluering efter implementering også, at man ikke kan randomisere – og dermed at man ikke kan få den metodisk bedste vurdering af effekten. Den ekstra udgift til randomisering, skal ses i forhold til hele udviklings- og implementeringsprojektet, og i forhold til det er randomiseringen sjældent den store post.

Andre metoder end lodtrækningsforsøg

Lodtrækningsforsøget er altså som nævnt i teorien den bedste metode til at besvare det kontrafaktiske spørgsmål – og altså dermed vurdere den kausale effekt. Men det er som påpeget ikke altid muligt at lave lodtrækningsforsøg. Men der findes selvfølgelig en række andre økonometriske metoder, som kan bruges til at estimere effekter, som fx IV, fixed effect, matching, eller diff-in-diff. Fælles for disse metoder er, at de på forskellig måde søger at løse det sektionsproblem, der opstår, når fordelingen mellem indsats- og kontrolgruppe ikke er tilfældig. Ud fra forskellige antagelser søger metoderne overordnet at korrigere for selektionsskævheden, således at outcomes for indsats- og kontrolgruppe kan sammenlignes – og derfor kalder vi også metoderne for quasi-eksperimentelle.

Der er selvfølgelig både fordele og ulemper ved de quasi-eksperimentelle metoder. Den altoverskyggende fordel er, at metoderne kan anvendes på allerede eksisterende data, og dermed kan bruges til at analysere effekter af indsatser tilbage i tiden – et retrospektivt design, hvor lodtrækningsforsøget er bundet til at foregå fremadrettet – et prospektivt design. Endvidere er de quasi-eksperimentelle metoder ikke så bundet af et bestemt set-up, men kan også drage nytte af data, der indsam-

les til andre formål. Omvendt er ulempen, at hver af metoderne har deres »blinde vinkler« i forhold til selektionsproblematikken, som afhængigt af kontekst kan udgøre et større eller mindre problem – eksempelvis bygger propensity score matching på matchning af observerbare karakteristika, mens der ikke mathes på uobserverbare karakteristika hos deltagere, og validiteten af en IV-estimator afhænger af det instrument, der anvendes. Endvidere er de indsamlede data ikke nødvendigvis de optimale i forhold til at analysere det ønskede spørgsmål. Et eksempel på, hvordan eksisterende data forsøges anvendt til at undersøge effekten af en given problemstilling, er Rangvid o.a. (2015), som på basis af registerdata søger at undersøge effekten af forberedende tilbud på unges overgang fra folkeskole til ungdomsuddannelse. Forfatterne afprøver forskellige metodiske designs, men er selvfølgelig begrænset af de data, der er til rådighed, og har derfor svært ved at konkludere noget meget præcist.

En anden ulempe ved de quasi-eksperimentelle metoder er, at kvalitetsvurderingen og sproget omkring den er mindre tydelig og veludviklet end for lodtrækningsforsøget. Dette skyldes dels, at der ikke var været samme tradition for kvalitetsvurdering af studier på dette område som for lodtrækningsforsøg, men også at det er utroligt metodisk komplekst at vurdere de forskellige metoder, og at én person aldrig kan være ekspert på samtlige studiedesigns.

Konsekvensen af dette er desværre, at det kan være svært – og næsten umuligt for lægmand – at vurdere kvaliteten af et givet studie. Og nogle af de eksempler på »talkrig«, som vi har set bryde ud i den offentlige debat – hvor den ene fløj siger »ja, det virker«, og den anden siger »nej, det gør ej« – skyldes netop at der er brugt forskellige metoder på forskellige datagrundlag, og at det er næsten umuligt at vurdere forskellene. Skal analyserne bruges i politisk sammenhæng (og det

er jo ofte målet), bør vi forskere derfor blive mere stringente til at fremlægge de metodiske overvejelser bag vores studie – for dermed at bidrage til at debatten kan blive både saglig og nuanceret.

Eksperimentelle forsøg i Danmark fremover

I de seneste 10 år er vi kommet langt i Danmark i forhold til at måle effekter af social- og velfærdspolitikken. Både på skoleområdet og socialområdet er der gang i flere lodtrækningsforsøg, fx i forhold til klasseledelse og forskellige forældretræningsprogrammer (f.eks. Keilow o.a., 2015; Christensen og Scavenius, 2015).

Men som beskrevet ovenfor, så er der også en del udfordringer. Det er udfordringer, som fint kan overkommes, men det kræver, at vi – dvs. forskere, beslutningstagere, interessenter og praktikere – bliver lidt mere ambitiøse i vores krav til viden om effekter. Vi skal være mere fokuserede på at tale om kausale effekter, vi skal stille flere krav til kvaliteten af de effektundersøgelser, der bliver lavet (og her er lodtrækningsforsøg som nævnt lettere at vurdere end andre metoder), og vi skal være mere målrettede og tålmodige i vores bestræbelser på at måle effekter.

Vi har i denne artikel talt meget om lodtrækningsforsøget. Det betyder dog ikke, at vi mener, at lodtrækningsforsøget er svaret på alt (hvilket nogen ellers vil hævde, hvis man har fulgt med i den meget polariserede evidensdebat på socialområdet). Lodtrækningsforsøg skal laves der, hvor man virkelig har brug for den mest solide viden om effekter og det kan lade sig gøre at lave god randomisering. Det kræver et godt samarbejde mellem forskning og praksis, men hvis man arbejder målrettet med det, er der ingen grund til at lade sig afskrække.

Litteratur

- Christensen, Christiane P. og Christoffer Scavenius (2015), »Et felteksperiment med Kærlighed i Kaos. Et forældretræningsprogram til familier med ADHD eller ADHD-lignende vanskeligheder«, København: Det Nationale Forskningscenter for Velfærd, 15:19.
- Christensen, Christiane P., Ida G. Andersen, Paul Bingley og Christoffer S. Sonne-Schmidt (2014), »Effekten af IT-støtte på elevers læsefærdigheder. Et felteksperiment i Horsens kommune«, København: Det Nationale Forskningscenter for Velfærd, 14:25.
- Christensen, Else (2000), »Det 3-årige barn. Rapport nr. 3 fra forløbsundersøgelsen af børn født i 1995«, København: Socialforskningsinstituttet, 00:10.
- Christensen, Else (2004), »7 års børneliv. Velfærd, sundhed og trivsel hos børn født i 1995«, København: Socialforskningsinstituttet, 04:13.
- Clinical Trials, <https://clinicaltrials.gov/ct2/home>
- CONSORT, www.consort-statement.org/
- Hansen, Helle og Maiken Pontoppidan (2014), »Erfaringer med RCT-studier på det sociale område«, Det Nationale Forskningscenter for Velfærd, notat.
- Higgins, Julian P.T. og Sally Green (2008), *Cochrane handbook for systematic reviews of interventions*, Chichester: Wiley-Blackwell.
- Keilow, Maria, Mette Friis-Hansen, Rune M. Kristensen og Anders Holm (2015), »Effekter af klasseledelse på elevers læring og trivsel«, København: Det Nationale Forskningscenter for Velfærd (kommende).
- Konnerup, Merete (2011), »Evidensbaseret politik«, *Samfundsøkonomen*, marts, nr 1.
- Mattson, Cathrine, Anne-Dorthe Hestbæk og Angelo R. Andersen (2008), »11-årige børns hverdagsliv og trivsel. Resultater fra SFT's forløbsundersøgelse af årgang 1995«, København: Det Nationale Forskningscenter for Velfærd, 08:16.
- Ottosen, Mai H., red. (2012), »15-åriges hverdagsliv og udfordringer. Rapport fra femte dataindsamling af forløbsundersøgelsen af børn født i 1995«, København: Det Nationale Forskningscenter for Velfærd, 12:30.
- Petrosino, A., C. Turpin-Petrosino, M. Hollos-Peel og J.G. Lavenberg (2013), »Scared straight and other juvenile awareness programs for preventing juvenile delinquency: A systematic review«, *Campbell Systematic Reviews* 2013:5.
- Phil, Brian, Anthony Biglan, Robert Boruch, Felipe Castro, Denise Gottfredson, Sheppard Kellam, Eve Moscicki, Steven Schinke, Jeffrey Valentine og Peter Ji (2005), »Standards of Evidence: Criteria for efficacy, effectiveness and dissemination«, *Prevention Science*, 6(3): 151-175.
- PRISMA, www.prisma-statement.org/
- Rangvid, Beatrice S., Vibeke M. Jensen og Sarah S. Nielsen (2015), »Forberedende tilbud og overgang til ungdomsuddannelse«, København: Det Nationale Forskningscenter for Velfærd, 15:14.
- Skolestøtte til børn i familiepleje (www.sfi.dk/skoleprojekt)
- Tanner-Smith, Emily E., Sandra J. Wilson, og Mark W. Lipsey (2013), »The comparative effectiveness of outpatient treatment for adolescent substance abuse: A meta-analysis«, *Journal of Substance Abuse Treatment*, 44(2): 145-58.
- Termansen, Tina og Christoffer S. Sonne-Schmidt (2012), »Forebyggende fysisk træning til ældre. En undersøgelse af effekten af en kort træningsindsats på ældres fysiske funktionsevne«, København: Det Nationale Forskningscenter for Velfærd, 12:33.
- Torgerson, David J. og Carole J. Torgerson (2008), *Designing randomised trials in health, education and the social sciences*, New York: Palgrave Macmillan.